Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)
Volume 5, No. 03 (2016), hal 311-320.

MODEL *STRATIFIED COX* UNTUK MENGATASI *NONPROPORTIONAL HAZARD*

Yesica Anastasya, Neva Satyahadewi, Hendra Perdana

INTISARI

Model *Stratified Cox* (SC) adalah modifikasi dari model *Cox Proportional Hazard* (PH) yang mengontrol *covariate* yang tidak memenuhi asumsi PH dengan membentuk strata, yaitu kombinasi dari semua kategori masing-masing *covariate* yang tidak memenuhi asumsi PH. Pembentukan model SC tersebut akan menghasilkan model *Cox* PH untuk setiap strata, sehingga variabel yang tidak memenuhi asumsi PH tetap dapat diamati efeknya dalam strata. Koefisien regresi dalam model diestimasi dengan memaksimumkan fungsi *partial likelihood,* dan kemudian diiterasi menggunakan metode Newton-Raphson. Data yang digunakan dalam penelitian ini adalah data *Criminal Recidivism* berupa waktu ditangkapnya kembali 432 mantan narapidana yang dipengaruhi oleh tujuh variabel yaitu bantuan dana (*Fin*), umur (*Age*), suku (*Race*), pengalaman bekerja (*Wexp*), status pernikahan (*Mar*), bebas bersyarat (*Paro*), tingkat pendidikan (*Educ*), dan banyak hukuman yang pernah diterima (*Prio*). Variabel yang tidak memenuhi asumsi PH adalah variabel *Age* yang merupakan variabel kontinu dengan rata-rata $24{,}59 \approx 25$ tahun yang kemudian akan dibentuk menjadi dua strata. Setelah dibandingkan berdasarkan nilai AIC/BIC, model SC tanpa interaksi lebih baik dibandingkan model *Cox* PH dan semua variabel dalam model SC telah memenuhi asumsi PH. Dapat disimpulkan, model SC dapat digunakan dalam mengatasi variabel yang tidak memenuhi asumsi PH dalam model *Cox* PH.

Kata Kunci: *Cox Proportional Hazard*, fungsi *survival*

PENDAHULUAN

Analisis *Cox Proportional Hazard* (PH) merupakan metode semiparametrik karena berbeda dengan model parametrik, dimana *baseline hazard* pada model parametrik mempunyai bentuk yang jelas mengikuti distribusi data *survival*. Dalam kenyataannya, data yang dimiliki tidak diketahui distribusinya, sehingga bentuk *baseline hazard* tidak diketahui. Walaupun distribusi data tidak diketahui, model Cox PH memberikan hasil yang mendekati distribusi data yang sebenarnya. [1]

Ketika model regresi Cox PH diterapkan pada data *survival*, hal yang harus dipenuhi adalah bahwa data harus memenuhi asumsi *proportional hazard*, yaitu *hazard ratio* untuk dua individu dengan nilai *covariate* yang berbeda konstan atau tidak berubah terhadap waktu. Apabila model tidak memenuhi asumsi PH atau biasa disebut *nonproportional hazard* maka ada variabel dalam model yang terikat oleh waktu sehingga variabel tersebut dikatakan "*time dependent variable*". *Nonproportional hazard* dapat menyebabkan tidak adanya perbedaan dalam interpretasi data akibatnya pemodelan regresi Cox PH pada model tersebut tidak tepat.

Pengecekan asumsi PH dilakukan dengan uji *Goodness of Fit* dan jika terdapat variabel yang *nonproportional* maka akan ditangani dengan menggunakan model regresi *Stratified Cox* (SC). Model SC adalah modifikasi dari model Cox PH yang memberikan perhatian atau mengontrol variabel yang tidak memenuhi asumsi PH tersebut dengan memungkinkan adanya "strata" atau tingkatan. Hal ini dilakukan karena diduga variabel yang tidak memenuhi asumsi PH tersebut tetap memiliki kontribusi, dan efeknya tetap diamati dengan menjadikannya strata yang tidak masuk ke dalam model.

Penelitian ini bertujuan untuk mengatasi variabel yang tidak memenuhi asumsi PH pada model *Cox* PH dengan membentuk model SC, serta penaksiran parameter pada model SC dengan metode *Partial Likelihood*.. Model SC memiliki dua tipe yaitu model SC tanpa interaksi dan model SC dengan interaksi.

Model SC tanpa interaksi mengasumsikan parameter yang sama untuk setiap strata, sedangkan model SC dengan interaksi mengasumsikan parameter yang berbeda untuk setiap strata. Model SC juga interaksi antar variabel dan variabel yang distratifikasi, sedangkan didalam model SC dengan interaksi terdapat interaksi antara variabel independen dengan variabel yang distratifikasi.

Data yang digunakan dalam penelitian ini berupa data *Criminal Recidivism* yang diteliti oleh Peter H. Rossi. Pada penelitiannya, Rossi membuat model regresi linear hubungan antara ditangkapnya kembali mantan narapidana dengan pemberian dana bantuan dana pasca bebas (TARP). Dalam penelitian ini, data *Criminal Recidivism* digunakan dalam analisis *survival* menghasilkan model *Cox* PH sehingga diketahui peluang mantan narapidana kembali ditangkap dipengaruhi oleh beberapa variabel independen. Langkah pertama dalam penelitian ini adalah menginput data *survival* dimana variabel dependen berupa waktu *survival* dan status tersensor subjek. Kemudian, setelah mengidentifikasi variabel dependen dan variabel independen dalam data, maka dibentuklah model Cox PH. Model *Cox* PH yang terbentuk kemudian melalui uji parameter untuk membuang variabel yang tidak signifikan dalam model. Berikutnya, dilakukan pengecekan asumsi PH menggunakan uji *Goodness of Fit*. Jika terdapat variabel yang tidak memenuhi asumsi PH, dibentuklah strata dan model Cox PH untuk masing-masing strata, yang kemudian disebut model *Stratified Cox* (SC). Setelah model SC tanpa interaksi dan model SC dengan interaksi dibuat, maka selanjutnya semua model yang dibuat dibandingkan berdasarkan nilai AIC/BIC untuk menentukan model terbaik.

DASAR TEORI ANALISIS SURVIVAL

Terdapat beberapa teori yang menjadi dasar dari analisis *survival*. Dasar-dasar teori tersebut antara lain adalah mengenai *probability density function* (pdf) dan *cumulative density function* (cdf), fungsi *survival* dan fungsi *hazard*. [2]

1. Fungsi Kepadatan Peluang

Fungsi kepadatan peluang $f(t)$, menyatakan probabilitas bahwa suatu peristiwa terjadi pada interval waktu $(t, t+\Delta t)$, sehingga dapat dituliskan sebagai:

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t}$$

Sehingga diberikan $F(t)$ merupakan *cdf* dari persamaan tersebut:

$$F(t) = P(T \leq t) = \int_0^t f(u)du$$

2. Fungsi *Survival*

Fungsi *survival* $S(t)$ didefinisikan sebagai peluang suatu individu dapat bertahan hidup dengan waktu *survival* sampai dengan waktu t dengan $(t > 0)$ yaitu:

$$S(t) = 1 - P(T \leq t) = 1 - F(t)$$

Secara matematis fungsi distribusi kumulatif dapat dinyatakan sebagai:

$$F(t) = \Pr(T \leq t) = 1 - \Pr(T > t) = 1 - S(t)$$

Fungsi *survival* juga dapat dinyatakan sebagai integral dari *probability density function* , $f(t)$, yaitu:

$$S(t) = \Pr(T > t) = \int_t^\infty f(u)du$$

$$f(t) = \frac{-dS(t)}{dt}$$

3. Fungsi Hazard

Fungsi *hazard* adalah probabilitas suatu individu mengalami *event* dalam interval waktu t sampai $t + \Delta t$ dengan syarat ia telah bertahan sampai waktu tersebut. Fungsi *hazard* yang dinotasikan $h(t)$ ialah fungsi yang menunjukkan rata-rata kegagalan saat waktu t, jika diketahui memiliki rata-rata waktu kegagalan pada interval $(t, \Delta t)$ dinyatakan dengan fungsi *hazard* berikut:

$$h(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t | \, T \geq t]}{\Delta t}$$

fungsi *hazard* dapat dinyatakan dalam bentuk berikut

$$h(t) = \frac{f(t)}{S(t)}$$

RESIDUAL SCHOENFELD

Residual *Schoenfeld* pada waktu kejadian *k* adalah [3]:

$$s_k = \int_{t_{k-i}}^{t_k} \sum_i \left[X_i - \bar{x}\left(\hat{\beta}, s\right)\right] dN_i(s) \tag{1}$$

Dengan

s_k : Nilai residual *Schoenfeld* pada kejadian ke *k*

X_i : Vektor *covariate* ke-*i*

$\bar{x}$: Rata-rata nilai *covariate*

N_i : Jumlah individu ke-*i* yang gagal saat waktu *t*

Himpunan residu *Schoenfeld* adalah matriks kolom yang berukuran *p* dengan satu baris untuk setiap kejadian. Ketika tidak ada kejadian yang berkaitan dengan waktu maka persamaan (2.3) dapat ditulis sebagai berikut:

$$s_k = X_{(k)} - \bar{x}(\hat{\beta}, t_k)$$

Dimana $X_{(k)}$ adalah vektor kovariat dari individu ke-*k* pada waktu yang sama. Residual *Schoenfeld* berguna untuk menaksir *proportional hazard*.

MODEL COX PROPORTIONAL HAZARD

Misalkan terdapat *p covariate* $\boldsymbol{X} = (X_1, X_2, \dots, X_p)$ dan nilai-nilai *covariate* yang dinyatakan dalam bentuk vektor **x** dimana $\boldsymbol{x} = (x_1, x_2, \dots., x_p)$, maka model Cox PH umum adalah sebagai berikut [4]:

$$h_i(t, \boldsymbol{X}) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) h_0(t)$$

Dengan

$h_i(t)$: Fungsi *hazard* untuk individu ke-*i*

$h_0(t)$: Fungsi *baseline hazard*

$\beta_1, \beta_2, \dots, \beta_p$: Koefisien regresi

$X_{i1}, X_{i2}, \dots, X_{ip}$: Variabel independen untuk individu ke-*i*

Hazard ratio didefinisikan sebagai *hazard* untuk satu individu dibagi dengan *hazard* untuk satu individu lain [1] :

$$\widehat{HR} = \frac{h_A(t, \boldsymbol{X}^*)}{h_B(t, \boldsymbol{X})} = \frac{h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i^{\,*})}{h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i)}$$

$$= \exp\left(\beta_i\left(\sum_{i=1}^{p} X_i{}^{*} - X_i\right)\right)$$

Dengan
$h_A(t, X^*)$: fungsi *hazard* untuk individu A
$h_B(t, X)$: fungsi *hazard* untuk individu B

Apabila nilai *hazard ratio* ($\widehat{HR}$) konstan sepanjang waktu, maka dapat dikatakan bahwa $X_1, X_2, \dots, X_p$ memenuhi asumsi PH.

UJI GOODNESS OF FIT

Uji *Goodness of Fit* untuk pengecekan asumsi PH adalah menguji korelasi antara *Schoenfeld residual* dengan rank *survival time* untuk masing-masing *covariate*. Pengujian asumsi PH terpenuhi untuk suatu *covariate* jika *Schoenfeld residual* untuk *covariate* tersebut tidak berkorelasi dengan waktu. [3]

Uji statistik yang digunakan meliputi:

a. Nilai *Rho* (*pearson product moment*): Korelasi antara residu *scaled Schoenfeld* dari tiap *covariate* dengan variabel transformasi waktu g_k.
b. Nilai *Chisq*: Nilai ini digunakan sebagai statistik uji, dirumuskan dengan:

$$T_j = \frac{\left\{\sum_k (g_k - \bar{g}) s_{kj}^*\right\}^2}{dI^{jj} \sum_k (g_k - \bar{g})^2}$$

Dengan:

g_k	: Variabel transformasi waktu
$\bar{g}$	: Rata-rata dari g_k
s_{kj}^*	: Residu *scaled Schoenfeld*
d	: Jumlah kematian/ kejadian
I^{jj}	: Diagonal matriks varians dari model Cox

MODEL STRATIFIED COX

Model *Stratified Cox* adalah modifikasi dari model *Cox* PH dimana dalam model *Stratified Cox* membagi fungsi *hazard* kedalam strata atau tingkatan-tingkatan dari *covariate*. *Covariate* yang dibagi kedalam strata ini merupakan *covariate* yang tidak memenuhi asumsi PH, yaitu dimana nilai $\widehat{HR}$ konstan terhadap waktu. [1]

Misalkan terdapat *m covariate*, model Cox PH yang terbentuk adalah

$$h(t, \boldsymbol{X}) = \exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \beta_{p+1} X_{p+1} + \cdots + \beta_m X_m) h_0(t) \quad (2)$$

Dari *m covariate* tersebut, dimisalkan terdapat *p covariate* yang memenuhi asumsi PH, dan *k covariate* yang tidak memenuhi asumsi PH, dengan $k = m - p$. Berdasarkan Persamaan (2), variabel-variabel yang tidak memenuhi asumsi PH adalah $X_{p+1}, X_{p+2}, \dots, X_m$, selanjutnya akan ditulis dalam notasi $Z_1, Z_2, \dots, Z_k$.

Tahap-tahap yang dilakukan dalam mengatasi variabel yang tidak memenuhi asumsi PH dalam model *Cox* PH yaitu:

a. Identifikasi variabel yang tidak memenuhi asumsi PH.
b. Definisikan variabel baru (Z^*) dengan mengkategorisasi variabel yang tidak memenuhi asumsi dan mengkombinasikan seluruh kategori setiap $Z_i, i = 1,2, .. k$. Z^* memiliki k^* kategori dengan k^* adalah total dari kombinasi (strata) hasil dari pengkategorian variabel-variabel Z_i.
c. Pembentukan model *Stratified Cox* tanpa interaksi dan model *Stratified Cox* dengan interaksi.

Model SC dibagi menjadi 2, yaitu:

1. Model SC Tanpa Interaksi

Model *Stratified Cox* tanpa interaksi adalah sebagai berikut:

$$h_g(t,\boldsymbol{X}) = h_{0g}(t)exp\left[\beta_1 X_1 + \beta_2 X_2 + \cdots . + \beta_p X_p\right]$$

g= 1,2,....,k*, jumlah strata dalam Z*

Variabel Z* tidak dimasukkan kedalam model, sedangkan $X_1, X_2, \ldots, X_p$ dimasukkan kedalam model. Untuk setiap strata (g= 1,2,....,k*) yang berbeda ,fungsi *baseline hazard* ($h_{0g}(t)$) dari setiap model berbeda. Sebaliknya, koefisien $\beta_1, \beta_2, \ldots., \beta_p$ sama disetiap strata.

2. Model SC dengan Interaksi

Berbeda dengan model SC tanpa interaksi, dalam membuat model SC dengan interaksi, maka kita mengasumsikan bahwa koefisien $\beta_1, \beta_2, \ldots., \beta_p$ berbeda disetiap strata. [1]

$$h_g(t,X) = h_{0g}(t)exp\left[\beta_{1g} X_1 + \beta_{2g} X_2 + \cdots + \beta_{pg} X_p\right]$$

$g = 1,2,\ldots,k^*$, jumlah strata dalam Z^*

Model SC dengan interaksi mengasumsikan adanya interaksi dari Z* variabel dengan variabel X sehingga model SC alternatif interaksi dapat dinyatakan dalam bentuk

$$h_g(t,X) = h_{0g}(t)exp\left[\beta_1 X_1 + \cdots + \beta_p X_p + \beta_{11}(Z_1^* \times X_1) + \cdots + \beta_{p1}\left(Z_1^* \times X_p\right) + \beta_{12}(Z_2^* \times X_1) + \cdots \right.$$
$$\left. + \beta_{p2}\left(Z_2^* \times X_p\right) + \cdots + \beta_{1,k^*-1}(Z_{k^*-1}^* \times X_1) + \cdots + \beta_{p,k^*-1}\left(Z_{k^*-1}^* \times X_p\right)\right]$$

$g = 1,2,\ldots,k^*$, jumlah strata dalam Z^*

dimana

$$Z_l^* = \begin{cases} 1\ , strata\ l+1 \\ 0, \qquad bukan \end{cases}, \quad l = 1, , \ldots, k^* - 1$$

ESTIMASI PARAMETER DENGAN PARTIAL LIKELIHOOD

Dimisalkan $X_j(t) = \left[X_{j1}(t), \ldots, X_{j1}(t)\right]^t$ adalah vektor *covariate* dari individu ke-*j* pada waktu *t*. Diberikan $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ adalah waktu terjadi *k* kali kegagalan dengan *covariate* $X_{(1)}, X_{(2)}, \ldots, X_{(k)}$. Diberikan juga $\boldsymbol{R}(t_{(i)})$ merupakan himpunan risiko (*risk set*) pada waktu $t_{(i)}$. Himpunan risiko ($\boldsymbol{R}(t_{(i)})$) terdiri dari semua individu yang waktu *survival*-nya paling sedikit $t_{(i)}$. Probabilitas kegagalan dari individu yang mengalami *event* pada saat $t_{(i)}$, dengan *covariate* $X_{(i)}$ dengan diketahui satu individu dari $\boldsymbol{R}(t_{(i)})$ yang gagal pada saat $t_{(i)}$, adalah:

$$\frac{\exp(\sum_{j=1}^{p} \beta_j X_{j(i)})}{\sum_{j \epsilon \boldsymbol{R}(t_{(i)})} \exp(\sum_{j=1}^{p} \beta_j X_{jl})} \tag{3}$$

Partial Likelihood dibentuk dengan mengalikan semua probabilitas bersyarat persamaan (3), untuk semua kematian atau kegagalan sehingga diperoleh fungsi *likelihood*-nya sebagai berikut [5] :

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp(\sum_{j=1}^{p} \beta_j X_{j(i)})}{\sum_{l \epsilon \boldsymbol{R}(t_{(i)})} \exp(\sum_{j=1}^{p} \beta_j X_{jl})}$$

Maximum partial likelihood estimator (MPLE) $\widehat{\boldsymbol{\beta}}$ dari $\boldsymbol{\beta}$ diperoleh dengan menyelesaikan persamaan Dengan $l(\boldsymbol{\beta}) = logL(\boldsymbol{\beta})$

$$\frac{\partial(l(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} = 0$$

UJI PARAMETER

Terdapat dua jenis uji dalam menguji parameter model, yaitu:

1. Pengujian secara serentak
   Pengujian secara serentak bertujuan untuk mengetahui apakah model signifikan atau tidak. Langkah-langkah uji hipotesis-nya adalah: [3]
   a. Uji hipotesis:
      $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$
      H_1: minimal satu $\beta_i \neq 0$ dengan $i = 1,2,\ldots,p$
   b. Tingkat signifikansi:
      $\alpha = 5\%$
   c. Statistik uji: $\chi^2_{LR} = 2[l(\boldsymbol{\beta}) - l(\beta_0)]$
   d. Daerah kritis:
      H_0 ditolak jika $\chi^2_{LR} > \chi^2_{\alpha;p}$ atau $\chi^2_{LR} \leq \chi^2_{(1-\alpha);p}$
   e. Kesimpulan:
      Jika H_0 ditolak, maka model secara umum adalah signifikan.
2. Pengujian secara parsial
   Untuk melihat variabel independen mana saja yang dapat signifikan mempengaruhi variabel dependen, dapat dilihat melalui *uji parsial.* Langkah-langkah uji hipotesisnya adalah:
   a. Uji hipotesis
      $H_0: \beta_i = 0 \quad i = 1,2,\ldots,p$ (Koefisien tidak layak masuk model)
      $H_1: \beta_i \neq 0 \quad i = 1,2,\ldots,p$ (Koefisien layak masuk model)
   b. Tingkat signifikansi:
      $\alpha = 5\% = 0.05$
   c. Statistik uji:
      $$Z = \frac{\hat{\beta} - \beta}{\sqrt{Var(\hat{\beta})}} \approx N(0,1)$$
      Dengan $Var(\hat{\beta}) = \frac{1}{n} E\left[\frac{\partial}{\partial \beta} \ln f(y_i|X;\hat{\beta})\right]^2$
   d. Darah kritik
      H_0 ditolak jika:
      Z_hitung > 1.96 atau Z_hitung < -1.96 atau $p.value < \alpha$
   e. Kesimpulan
      Jika H_0 ditolak, maka koefisien layak masuk dalam model.

STUDI KASUS

Data yang memiliki variabel yang tidak memenuhi asumsi PH diatasi dengan model membentuk *Stratified Cox* dengan data yang digunakan ialah data *Criminal Recidivism* yang diteliti oleh Peter H. Rossi pada tahun 1976 di Texas, Amerika Serikat. Data mempunyai 6 variabel independen kategorik yaitu *Fin (Financial Aid), Race, Wexp (Work experience), Mar (Marriage), Paro (Parole), Educ (Education), Prio* dan 1 variabel independen kontinu yaitu *Age*. Model umum Cox PH yang akan dibuat berdasarkan data adalah sebagai berikut

Tabel 1 Model regresi *Cox* data *Criminal Recidivism*

Variabel	Coef	Exp(Coef)	se(Coef)	P-value
Age	-0,0512	0,950	0,022	0,021
Prio	0,0791	1,082	0,029	0,007
Educ 3	0,5912	1,810	0,519	0,255

Tabel 1 Model regresi *Cox* data *Criminal Recidivism (Lanjutan)*

Variabel	**Coef**	**Exp(Coef)**	**Se(Coef)**	**P-Value**
Educ 4	0,3267	1,390	0,544	0,548
Educ 5	-0,1208	0,886	0,675	0,858
Educ 6	-0,4066	0,665	1,123	0,717
Mar	-0,4229	0,655	0,382	0,268
Fin	-0,4021	0,669	0,193	0,037
Race	0,3616	1,436	0,312	0,247
Wexp	-0,1215	0,886	0,213	0,569
Paro	-0,0982	0,906	0,196	0,616

Dari Tabel 1, model regresi *Cox* awal untuk data *Criminal Recidivism* yaitu:

$$\begin{aligned} h(t) = h_0(t)\exp[&-0{,}0512\,Age + 0{,}0791 Prio + 0{,}5912 Educ\,3 + 0{,}3267\,Educ\,4 \\ &- 0{,}1208 Educ\,5 - 0{,}4066 Mar - 0{,}4021\,Fin + 0{,}3616\,Race \\ &- 0{,}1215 Wexp - 0{,}0982 Paro] \end{aligned}$$

Selanjutnya, variabel-variabel akan diseleksi dengan metode backward. Seleksi variabel dapat dilakukan dengan ukuran kriteria AIC. Dengan metode *backward*, setiap variabel independen yang tidak signifikan dan memiliki nilai AIC terkecil akan dihapus dari model. Proses ini dilanjutkan sampai kita mendapatkan model regresi *Cox* terbaik dengan nilai AIC model terkecil.

Dengan metode *backward* diperoleh model terbaik dengan AIC terkecil dengan variabel *Age, Prio, Mar,* dan *Fin* yang akan membentuk model *Cox* terbaik, yaitu:

Tabel 2 Model regresi *Cox* terbaik

Variabel	**Coef**	**Exp(Coef)**	**se(Coef)**	**z**	**P-value**
Age	-0,060	0,940	0,021	-2,891	0,0038
Prio	0,097	1,102	0,027	3,571	0,0004
Mar	-0,532	0,587	0,372	-1,428	0,1531
Fin	-0,359	0,697	0,190	-1,889	0,0589

Dari Tabel 2 model *Cox* terbaik data *Criminal Recidivism* adalah sebagai berikut:

$$h(t) = h_0(t)\exp[-0{,}06Age + 0{,}097Prio - 0{,}532\,Mar - 0{,}359Fin\,] \tag{4}$$

Selanjutnya akan dilakukan uji asumsi PH dari model (4) menggunakan uji *Goodness of Fit* (*Gof*).

Tabel 3 Uji asumsi PH dengan *Gof*

	Rho	***Chisq***	***P-Value***
Age	-0,2201	7,342	0,00675
Prio	-0,0770	0,719	0,39645
Mar	0,1310	2,060	0,15077
Fin	-0,0002	$2{,}8\times10^{-6}$	0,99865

Dari Tabel 3, terlihat bahwa nilai *P-value* untuk variabel *Age* kurang dari 5%, sedangkan untuk variabel lainnya nilai *p.value* lebih dari 5%, sehingga dapat disimpulkan bahwa variabel *Age* tidak memenuhi asumsi PH.

Variabel *Age* merupakan variabel kontinu sehingga harus dikategorisasi terlebih dahulu. Variabel ini akan dikategorikan menjadi 2 kategori yaitu umur diatas rata-rata dan umur dibawah rata-rata, dengan rata-rata umur 24,6 ≈ 25 tahun, sehingga hasilnya adalah $Age = \begin{cases} 1, jika\ Age \le 25 \\ 0, jika\ Age > 25 \end{cases}$

Setelah variabel *Age* dikategorisasi, variabel *Age* akan dikeluarkan dari model dan menjadi strata, sedangkan variabel lainnya tetap masuk didalam model. Selanjutnya dibentuk model sebagai berikut:

1. Model SC Tanpa Interaksi

Fungsi hazard dalam model SC tanpa interaksi untuk data *Criminal Recidivism* adalah sebagai berikut:

$$h_g(t,\boldsymbol{X}) = h_{0g}(t)\exp[\beta_1 Fin + \beta_2 Prio + \beta_3 Mar] \quad (5)$$

Dimana: g : 1, 2

$h_{0g}(t)$: *baseline hazard* untuk strata g

Setelah estimasi parameter diperoleh model SC tanpa interaksi sebagai berikut

Tabel 4 Model *Stratified Cox* tanpa interaksi

Variabel	Coef	Exp(Coef)	se(Coef)	z	P-value
Prio	0,098	1,104	0,027	3,687	0,0002
Mar	-0,645	0,525	0,370	-1,741	0,0816
Fin	-0,393	0,680	0,190	-2,064	0,0390

Dari Tabel 4, model SC tanpa interaksi yang terbentuk adalah sebagai berikut:

$$h_g(t,\boldsymbol{X}) = h_{0g}(t)\exp[-0{,}393 Fin + 0{,}098 Prio - 0{,}645 Mar]$$

Berikutnya dilakukan pengecekan asumsi PH untuk masing-masing variabel yang ada dalam model SC tanpa interaksi. Hasil dari uji *Goodness of Fit* adalah sebagai berikut:

Tabel 5 Uji asumsi PH pada model SC tanpa interaksi

	Rho	*Chisq*	*P-Value*
Prio	-0,0890	0,9278	0,335
Mar	-0,0035	0,0015	0,970
Fin	0,1204	1,6852	0,194

Dari Tabel 5, semua variabel independen dalam model SC tanpa interaksi memiliki nilai *P-value* lebih besar dari 0,05 (*P-value* > 0,05), sehingga dapat disimpulkan model SC tanpa interaksi memenuhi asumsi PH.

2. Model SC dengan Interaksi

Model SC dikatakan model SC dengan interaksi karena bahwa koefisien $\beta_1, \beta_2, \ldots, \beta_p$ berbeda disetiap strata. Berikut adalah model *Cox* PH untuk masing-masing kategori umur:

Tabel 6 Model *Cox* PH untuk kategori *Age=1*

Variabel	Coef	Exp(Coef)	se(Coef)	z	P-value
Prio	0,1004	1,1056	0,0303	3,310	0,0009
Fin	-0,3158	0,7290	0,2152	-1,467	0,1423
Mar	-0,3670	0,6932	0,4236	-0,865	0,3869

Tabel 7 Model *Cox* PH untuk kategori *Age=0*

Variabel	Coef	Exp(Coef)	se(Coef)	z	P-value
Prio	0,0861	1,089	0,0553	1,557	0,1194
Fin	-0,6521	0,5209	0,4086	-1,596	0,1106
Mar	-1,2350	0,2908	0,7380	-1,673	0,0943

Dari Tabel 6 dan Tabel 7, model SC dengan interaksi yang terbentuk yaitu:

$$h_g(t,X) = h_{0g}(t)exp\left[\beta_{1g} Prio + \beta_{2g} Fin + \beta_{3g} Mar\right]$$

Dimana $g = 1$ (Umur dibawah 25) dan $g = 2$ (Umur diatas 25 tahun)

$Age \geq 25$: $h_1(t,X) = h_{01}(t)exp[\beta_{11} Prio + \beta_{21} Fin + \beta_{31} Mar]$

$h_1(t,X) = h_{01}(t)exp[0{,}1004 Prio - 0{,}3158 Fin - 0{,}367 Mar]$

$Age < 25$: $h_2(t,X) = h_{02}(t)exp[\beta_{12} Prio + \beta_{22} Fin + \beta_{32} Mar]$

$$h_2(t,X) = h_{02}(t)exp[0{,}0861Prio - 0{,}6521Fin - 1{,}235Mar]$$

Model SC alternatif interaksi mengasumsikan adanya interaksi dari Z^* variabel dengan variabel X, yang dapat dinyatakan sebagai berikut:

Tabel 8 Model *Stratified Cox* dengan Interaksi

Variabel	Coef	Exp(Coef)	se(Coef)	z	P-value
Prio	0,086	1,089	0,055	1,557	0,119
Fin	-0,652	0,521	0,408	-1,590	0,110
Mar	-1,235	0,291	0,738	-1,670	0,094
Age.prio	0,014	1,014	0,063	0,226	0,821
Age.fin	0,336	1,399	0,462	0,728	0,466
Age.mar	0,868	2,383	0,851	1,021	0,307

Dari Tabel 8, model SC alternatif interaksi yaitu:

$$h_g(t,X) = h_{0g}(t)\exp[0{,}086\,Prio - 0{,}652\,Fin - 1{,}235\,Mar + 0{,}014\ \ (Age \times Prio) + 0{,}336\,(Age{\times}Fin) + 0{,}868(Age{\times}Mar)]$$

Jika disubstitusikan nilai *Age* sesuai kategori $Age = \begin{cases} 1, jika\ Age \leq 25 \\ 0, jika\ Age > 25 \end{cases}$, model alternatif interaksi akan menghasilkan bentuk umum model SC dengan interaksi

Berikutnya dilakukan pengecekan asumsi PH untuk masing-masing variabel yang ada dalam model SC dengan interaksi. Hasil dari uji *Goodness of Fit* adalah sebagai berikut

Tabel 9 Uji asumsi PH pada model SC dengan interaksi

	Rho	***Chisq***	***P-Value***
Prio	0,1050	1,344	0,2463
Mar	-0,0586	0,398	0,5280
Fin	0,1579	2,823	0,0929
Age.Prio	-0,1706	3,497	0,0615
Age.Mar	-0,1215	1,669	0,9164
Age.Fin	0,0589	0,401	0,5266

Dari Tabel 9, semua variabel independen dalam model SC dengan interaksi memiliki nilai *P-value* lebih besar dari 0,05 (*P-value* > 0,05), sehingga dapat disimpulkan model SC dengan interaksi memenuhi asumsi PH.

Tabel perbandingan ketiga model tersebut adalah sebagai berikut:

Tabel 10 Perbandingan nilai AIC dan BIC berdasarkan model

MODEL	AIC	BIC
Model *Cox* PH	1334,86	1364,961
Model *Cox* PH (tanpa variabel *Age*)	1328,10	1339,056
Model SC tanpa interaksi	1214,31	1222,515
Model SC dengan Interaksi	1218,61	1235,029

Berdasarkan Tabel 10, model *Stratified Cox* tanpa interaksi memiliki nilai AIC dan BIC minimum dibandingkan model lainnya, yaitu dengan AIC=1214,307 dan BIC=1222,515. Oleh karena itu, dapat disimpulkan bahwa model *Stratified Cox* tanpa interaksi merupakan model terbaik untuk data *Criminal Recidivism* dengan interpretasi model yaitu

a. Mantan narapidana yang menerima bantuan dana pasca bebas (*Fin*=1) memiliki risiko ditangkap kembali 0,68 kali lebih kecil dibandingkan yang tidak menerima bantuan dana.
b. Bertambahnya jumlah hukuman mantan narapidana (*Prio*) sebanyak 1 hukuman, bertambah juga risiko ditangkap kembali sebesar 1,104 kali.

c. Risiko ditangkap kembali mantan narapidana yang sudah menikah (*Mar*=1) 0,525 kali lebih kecil dibandingkan yang tidak menikah.

PENUTUP

Berdasarkan hasil analisis dan pembahasan, maka diperoleh kesimpulan bahwa model *Stratified Cox* dapat digunakan untuk mengatasi model *Cox* PH yang memiliki variabel yang *nonproportional hazard,* yaitu dengan membentuk "strata", yaitu kombinasi dari semua kategori masing-masing variabel yang tidak memenuhi asumsi PH. Dalam penerapannya pada data *Criminal Recidivism,* variabel yang tidak memenuhi asumsi PH adalah variabel *Age* yang merupakan variabel kontinu. Variabel dibagi menjadi 2 strata, kemudian dibentuk model SC tanpa interaksi dan model SC dengan interaksi. Dengan uji *Goodness of Fit* seluruh variabel dalam kedua model SC memiliki nilai $P - value > 0{,}05$, sehingga kedua model SC telah memenuhi asumsi PH. Model *Stratified Cox* tanpa interaksi memiliki nilai AIC dan BIC minimum dibandingkan model lainnya, yaitu dengan AIC=1214,307 dan BIC=1222,515. Oleh karena itu, dapat disimpulkan bahwa model *Stratified Cox* tanpa interaksi merupakan model terbaik untuk data *Criminal Recidivism*

DAFTAR PUSTAKA

[1] Kleinbaum, D.G. dan Klein, M. *Survival Analysis- A Self Learning Text, Second Edition,* New York: Springer; 2005.
[2] Klein, J.P dan Moeschberger, M.L. *Survival Analysis Techniques for Censored and Truncated Data*. New York: Springer; 2003.
[3] Therneau, T.M dan Grambsch, P.M. *Modeling Survival Data:Extending the Cox Model.* New York: Springer; 2000.
[4] Collet, D. *Modelling Survival Data in Medical Research.* London: Chapman and Hall; 1994
[5] Lee, E.T. dan Wang, J.W. *Statistical Methods for Survival Data Analysis (Third Edition)*. New York: John Willey and Sons Inc; 2003.

YESICA ANASTASYA : Fakultas MIPA Universitas Tanjungpura, Pontianak, yesica.anastasya2012@gmail.com
NEVA SATYAHADEWI : Fakultas MIPA Universitas Tanjungpura, Pontianak, neva.satya@math.untan.ac.id
HENDRA PERDANA : Fakultas MIPA Universitas Tanjungpura, Pontianak, hendra.perdana@math.untan.id